

Steekproefgrootte van een studie

Rapportering van de steekproefgrootte in gepubliceerde studies

P. Chevalier

Het doel van een interventiestudie is na te gaan of een gegeven behandeling een klinisch relevant effect heeft in vergelijking met een referentiebehandeling of met placebo. Om statistisch betrouwbare resultaten te bekomen, moet de steekproef van de studie bestaan uit voldoende personen die de interventie ondergaan.

Steekproefberekening van een studie

De conventionele aanpak voor de berekening van de nodige steekproefgrootte berust op vier elementen¹: type-I-fout, power van de studie (type-II-fout), verwachte respons (met standaarddeviatie) in de controlegroep en tenslotte het te verwachten effect van de behandeling.

Een type-I-fout (α -fout) in een studie waarbij men de superioriteit van één behandeling ten opzichte van een andere wil aantonen, is het onterecht verwerpen van de nulhypothese, m.a.w. besluiten dat er een verschil is tussen de twee behandelingen, terwijl er toch geen verschil is. Algemeen wordt een risico van 5% geaccepteerd, wat overeenkomt met een p-waarde van 0,05. Een type-II-fout (β -fout) is binnen dezelfde context onterecht de nulhypothese accepteren, m.a.w. besluiten dat er geen verschil bestaat tussen twee behandelingen, terwijl dit in werkelijkheid wel het geval is. De term 'power' geeft aan hoeveel patiënten er moeten geïncludeerd worden om een type-II-fout te vermijden. Over het algemeen wordt een power gekozen van minstens 80%. De schatting van het aantal gebeurtenissen in de controlegroep berust vaak op een schatting in een pilootstudie bij de beoogde patiëntenpopulatie. Voor de berekening van de steekproef gebruikt men de variantie van de resultaten als aanwijzing voor de spreiding van de resultaten. Zich baseren op een pilootstudie staat momenteel ter discussie, omdat deze methode minder betrouwbaar is. Verschillende alternatieven zijn mogelijk. Men kan zich baseren op het minimale klinische belangrijke verschil of men kan de steekproefgrootte zodanig berekenen dat de verwachte breedte van het betrouwbaarheidsinterval een bepaalde vooraf vastgelegde waarde niet overschrijdt². Het te verwachten effect van een behandeling is het minimale detecteerbare effect tussen twee behandelingen dat klinisch voldoende groot en relevant is. Beide laatste elementen zijn het moeilijkst vast te leggen en kunnen voor gevolg hebben dat de studie onvoldoende power heeft. Op basis van deze vier elementen kan de steekproefgrootte berekend worden met behulp van een wiskundige formule² of van een nomogram, ontwikkeld door Core en Altman³.

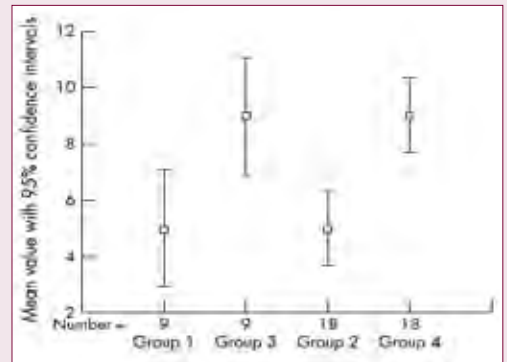
Een steekproefgrootte post-hoc berekenen is voor de kliniek niet nuttig. Na het uitvoeren van de studie geven de betrouwbaarheidsintervallen de meest relevante informatie². De breedte van het betrouwbaarheidsinterval is omgekeerd evenredig aan het aantal geïncludeerde patiënten. Een kleiner betrouwbaarheidsinterval verhoogt de nauwkeurigheid van het resultaat, wat soms bepalend is om een significant verschil aan te tonen, dat niet kan aangetoond worden bij een te kleine steekproef.

Illustratie (herwerkt op basis van referentie 3)

Resultaten van de evaluatie van een behandeling in vier verschillende groepen

Groepen 1 en 3: negen deelnemers in elke groep, met gedeeltelijk verschillende resultaten in de twee groepen, dus verschillende gemiddelden, maar met overlapping van de betrouwbaarheidsintervallen, wat betekent dat het verschil tussen beide groepen statistisch niet significant is. Groepen 2 en 4: 18 deelnemers in elke groep; de deelnemers behalen dezelfde gemiddelden in groep 2 versus groep 1 en in groep 4 versus groep 3, maar de betrouwbaarheidsintervallen tussen groepen 2 en 4 overlappen niet meer, wat wijst op een statistisch significant verschil tussen beide groepen.

Dit voorbeeld toont aan dat door het steekproefaantal te verhogen, de overlapping (= verschil kan aan toeval te wijten zijn) van de betrouwbaarheidsintervallen in de 2 groepen verdwijnt en dat er een statistisch significant verschil kan aangetoond worden.



Rapportering van de steekproefberekening in gepubliceerde studies

Is het mogelijk om in de publicatie van een studie na te gaan of de steekproefgrootte correct werd berekend bij het uitwerken van het protocol? Charles et al. onderzochten in 2009 de kwaliteit van de rapportering van de steekproefberekening in gerandomiseerde studies met een superioriteitsprotocol¹. De auteurs includeerden RCT's met één enkele primaire uitkomstmaat, gepubliceerd in zes medische tijdschriften (2005-2006) met hoge impactfactor. De resultaten waren niet erg geruststellend. Van de 215 geselecteerde publicaties vermeldde 5% geen enkele steekproefberekening en 43% gaf geen informatie over alle vereiste parameters. Charles et al. hadden voor 157 studies (70%) voldoende gegevens om opnieuw de steekproefgrootte te berekenen. Het verschil tussen de steekproefgrootte vermeld in het artikel en de herberekende steekproefgrootte was groter dan 10% in 47 studies (30%). Het verschil tussen de vooronderstellingen voor de controlegroep en de geobserveerde data was groter dan 30% in 45 publicaties (31%) en meer dan 50% in 24 publicaties (17%). Slechts 34% van de studies vermeldde alle gegevens die nodig zijn voor het berekenen van de steekproef, voerde een accurate berekening uit van de steekproef en schatte exact het verloop in van de beoogde uitkomst in de controlegroep. De auteurs besluiten dat er vragen gesteld kunnen worden bij de manier waarop men in RCT's steekproeven berekent.

Besluit

Er bestaan duidelijke regels voor het berekenen van de steekproefgrootte in een studie. Het is echter onmogelijk om voor de meeste RCT's, gepubliceerd in gerenommeerde tijdschriften, na te gaan of deze regels goed zijn toegepast.

Referenties

1. Charles P, Giraudeau B, Dechartres A, et al. Reporting of sample size calculation in randomised controlled trials: review. *BMJ* 2009;338:b1732.
2. Brasher PMA, Brant FB. Sample size calculations in randomized trials: common pitfalls. *Can J Anesth* 2007;54:103-6.
3. Jones SR, Carley S, Harrison M. An introduction to power and sample size estimation. *Emerg Med J* 2003;20:453-8.